

CS 785: Theoretical Computer Science Toolkit

Lecture 5: Chernoff Bounds and Concentration Inequalities

Instructor: Sruthi Sekar
Scribe: Abhishek Kumar
Roll no:24M0802

January 28, 2025

Lecture Overview

This lecture presents a comprehensive treatment of Chernoff bounds, starting from classical inequalities (Markov, Chebyshev), going through the second moment method, and introducing the powerful method of bounding moment generating functions. We explore tight exponential bounds on tail probabilities of sums of independent random variables, with generalizations like Hoeffding's, Bernstein's, and McDiarmid's inequalities. Complete proofs, derivations, and motivating examples are provided, along with a visualization of the Chernoff bound idea.

1 Motivation: Why Tail Bounds Matter

Let $X_1, X_2, \dots, X_n$ be i.i.d. real-valued random variables with mean μ and variance σ^2 , and define:

$$S_n = \sum_{i=1}^n X_i$$

By the Central Limit Theorem (CLT),

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1)$$

However, CLT doesn't tell us how likely it is that S_n deviates significantly from $\mathbb{E}[S_n]$.

Berry–Esseen Theorem

It quantifies the convergence in CLT:

$$\left| \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq t\right) - \Phi(t) \right| \leq \frac{C \cdot \rho}{\sigma^3\sqrt{n}}$$

where $\rho = \mathbb{E}[|X_i - \mu|^3]$. But the error decays slowly: $O(1/\sqrt{n})$.

2 Markov's Inequality

For $X \geq 0$:

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$$

Proof: Let $A = \{X \geq t\}$. Then $X \geq t \cdot \mathbf{1}_A$ implies:

$$\mathbb{E}[X] \geq t \cdot \mathbb{P}[X \geq t] \Rightarrow \mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$$

3 Chebyshev's Inequality

If $\mathbb{E}[X] = \mu$, $\text{Var}(X) = \sigma^2$:

$$\mathbb{P}[|X - \mu| \geq t] \leq \frac{\sigma^2}{t^2}$$

Proof: Apply Markov's inequality to $(X - \mu)^2$:

$$\mathbb{P}[(X - \mu)^2 \geq t^2] \leq \frac{\mathbb{E}[(X - \mu)^2]}{t^2} = \frac{\sigma^2}{t^2}$$

4 Second Moment Method

Let $X \geq 0$, $\mathbb{E}[X] = \mu > 0$, $\mathbb{E}[X^2] = \nu$. Then:

$$\mathbb{P}[X > 0] \geq \frac{\mu^2}{\nu}$$

Proof: Use Cauchy-Schwarz on $\mathbb{E}[XI]$ and $\mathbb{E}[I^2]$ where $I = \mathbf{1}_{X>0}$.

5 Example: Rademacher Variables

Let $X_i \in \{-1, +1\}$ be i.i.d. Rademacher variables. Then $X = \sum X_i$:

$$\begin{aligned}\mathbb{E}[X] &= 0 \\ \mathbb{E}[X^2] &= n \\ \mathbb{E}[X^4] &= 3n^2\end{aligned}$$

Using fourth moment bound:

$$\mathbb{P}[|X| \geq t] \leq \frac{3n^2}{t^4}$$

6 Moment Generating Functions and Tail Bounds

Let X be real-valued. For $\lambda > 0$:

$$\mathbb{P}[X \geq t] \leq e^{-\lambda t} \mathbb{E}[e^{\lambda X}]$$

If $X = \sum X_i$ with independent X_i :

$$\mathbb{E}[e^{\lambda X}] = \prod \mathbb{E}[e^{\lambda X_i}]$$

This leads to Chernoff bounds.

7 Chernoff Bound: Full Derivation

Let $X_1, \dots, X_n \in [0, 1]$ independent, $S = \sum X_i$, $\mu = \mathbb{E}[S]$. We want:

$$\mathbb{P}[S \geq (1 + \delta)\mu] \leq ?$$

By Markov:

$$\mathbb{P}[S \geq (1 + \delta)\mu] \leq \inf_{\lambda > 0} \frac{\mathbb{E}[e^{\lambda S}]}{e^{\lambda(1+\delta)\mu}}$$

Using Hoeffding's lemma:

$$\mathbb{E}[e^{\lambda X_i}] \leq e^{\lambda^2/8} \text{ (for } X_i \in [0, 1])$$

Then:

$$\mathbb{E}[e^{\lambda S}] \leq \exp\left(\lambda\mu + \frac{n\lambda^2}{8}\right)$$

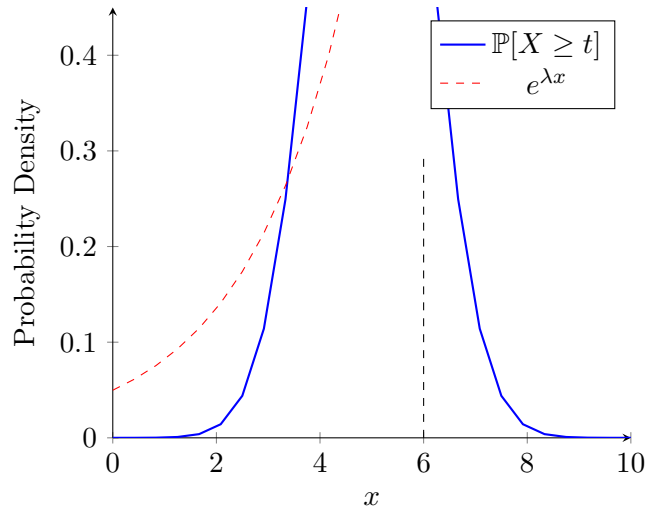
Thus:

$$\mathbb{P}[S \geq (1 + \delta)\mu] \leq \exp\left(-\lambda(1 + \delta)\mu + \lambda\mu + \frac{n\lambda^2}{8}\right)$$

Choose $\lambda = \frac{4\delta\mu}{n}$ to minimize. This gives:

$$\mathbb{P}[S \geq (1 + \delta)\mu] \leq \exp\left(-\frac{\delta^2\mu}{3}\right)$$

8 Chernoff Bound Diagram



9 Hoeffding's Inequality

Let $X_i \in [a_i, b_i]$. Then:

$$\mathbb{P}[S - \mathbb{E}[S] \geq t] \leq \exp\left(-\frac{2t^2}{\sum (b_i - a_i)^2}\right)$$

Proof: Apply MGF bounds and Hoeffding's lemma.

10 Bernstein's Inequality

Let X_i be zero-mean, $|X_i| \leq M$, $\text{Var}(S) = \sigma^2$. Then:

$$\mathbb{P}[S \geq t] \leq \exp\left(-\frac{t^2}{2\sigma^2 + 2Mt/3}\right)$$

Useful when variance is small compared to the maximum value.

11 McDiarmid's Inequality

Let $f : \mathcal{X}^n \rightarrow \mathbb{R}$ with:

$$|f(x_1, \dots, x_n) - f(x_1, \dots, x'_i, \dots, x_n)| \leq c_i$$

Then:

$$\mathbb{P}[f(X) - \mathbb{E}[f(X)] \geq \epsilon] \leq \exp\left(-\frac{2\epsilon^2}{\sum c_i^2}\right)$$

Proof: Construct a Doob martingale and apply Azuma's inequality.

12 Sampling Corollary

Let $X_1, \dots, X_n \in [0, 1]$ i.i.d., $\mu = \mathbb{E}[X_i]$, $\hat{\mu} = \frac{1}{n} \sum X_i$:

$$\mathbb{P}[|\hat{\mu} - \mu| \geq \epsilon] \leq 2 \exp(-2n\epsilon^2)$$

Proof: Apply McDiarmid's inequality with $c_i = 1/n$.